

# ON TEMPLE–KATO LIKE INEQUALITIES AND APPLICATIONS

LUKA GRUBIŠIĆ

**ABSTRACT.** We give both lower and upper estimates for eigenvalues of unbounded positive definite operators in an arbitrary Hilbert space. We show scaling robust relative eigenvalue estimates for these operators in analogy to such estimates of current interest in Numerical Linear Algebra. Only simple matrix theoretic tools like Schur complements have been used. As prototypes for the strength of our method we discuss a singularly perturbed Schroedinger operator and study convergence estimates for finite element approximations. The estimates can be viewed as a natural quadratic form version of the celebrated Temple–Kato inequality.

## 1. INTRODUCTION

The purpose of this article is to establish scaling robust estimates for discrete eigenvalues of positive definite operators in a Hilbert space. We also prove that our estimates are optimal for a residual type analysis of lowermost eigenvalues of those operators. Our approach uses the theory of quadratic forms from [21, Chapters VI–VIII] and an adaptation of the matrix relative perturbation theory. As a result we establish the same high performance residual type estimates from [11] in our more general setting. For a review of the matrix relative perturbation theory see [24] and the references therein.

It turns out that positive-definiteness of the matrix is the key structural property which is needed for the analysis of [11, 12]. Subsequently, we prove our estimates for an abstract positive definite form in an arbitrary Hilbert space. This, together with the fact that the estimates also hold for the discrete eigenvalues which are in gaps of the essential spectrum, indicates that our simple matrix analytic techniques are well adapted to the class of problems under study, e.g. our technique yields high performance estimates without forcing us to impose any unnecessary restrictions. This abstract approach is further justified by the fact that we simultaneously consider applications of these estimates to a study of the convergence properties of adaptive finite element methods as well as to a quantitative study of the asymptotic properties of eigenvalue problems in the large coupling limit. Typical operators in the large coupling limit setting are those from [3, 8, 29].

The obtained estimates are the same as those which have proven themselves in [12] as a significant tool in the development of modern mathematical software. The main feature of the matrix eigenvalue algorithms from [12] is that they are robust when applied to extremely badly scaled input matrices. We bring this in correspondence with the behavior of the spectrum of stiffly/singularly perturbed operators from [3, 8, 29]. We use a model Schroedinger operator to

---

*Key words and phrases.* Estimation of eigenvalues, upper and lower bounds, Eigenvalues, Variational methods for eigenvalues of operators.

This work is based on a part of author’s PhD thesis [16], which was written under the supervision of Prof. Dr. Krešimir Veselić, Hagen in partial fulfilment of the requirements for the degree Dr. rer. nat.

show that our estimates are optimal for the class of stiff/singular perturbations, see Section 4. An extensive study of non-inhibited stiff families of operators has been performed, with the help of the results from this article, in [16] and will be published in the subsequent report.

When studying the convergence of finite element procedures we adopt the approach of [1, 25]. Our results give a new flavor to the analysis of the stopping criteria for preconditioned inverse iterations from [25, Section 4]. In comparison, we are more explicit about the dependence of the ingredients of the error on the input data and we can prove the equivalence of the error and the estimator, see Section 5. Furthermore, in the Conclusion we briefly outline a simple way to obtain optimal eigenvector estimates from our eigenvalue results.

In Section 1.1 we relate our new results—which appear in Section 3—to other results in the literature. In Section 1.2 we use a simple matrix model problem to give a first flavor of the results by comparing our approach with that of the Temple–Kato inequality.

**1.1. A comparison with other approaches.** A number of recent studies of the eigenvalue approximation problem—when classified from the viewpoint of perturbation theory (see e.g. [19, 24])—could be seen to fall into the following classes:

- (1) Results obtained in the *absolute* (sometimes also referred to as regular) setting. For recent results see [26] and references therein. The performance of such estimates, when applied to an unbounded operator, depends on the method of the regularization. The delicacy of this issue is illustrated on an example below.
- (2) Results which are obtained by interpreting the eigenvalue problem as a nonlinear problem (in both eigenvalue and eigenvector). This approach makes a treatment of the eigenvalue multiplicity somewhat more difficult, since it is not easy to profit from the special structure which the eigenvalue problem has, cf. [18, Remark 7].
- (3) Direct analysis of a representation of the single vector *residual*  $r = \mathbf{H}\psi - \mu\psi$ , coupled with a consideration of the approximation properties of the function space which is used to generate the *test vector*  $\psi$ ,  $\|\psi\| = 1$ , where we have used  $\mu := (\psi, \mathbf{H}\psi)$  to denote the *Ritz value*. Such estimates are essentially asymptotic in nature (cf. [13, Remark 3.2]) or are specifically tailored for the particular (class of) problem(s) under study (cf. [23, Theorem 4.1]).

We propose a technique which is based on the *relative* perturbation theory for quadratic forms in a Hilbert space from [21, Chapters VI–VIII]. In this regularization framework (which was developed in [15, 17]) we solely use elementary matrix techniques, like LU-decomposition and Schur complements, to obtain new block operator residual equation. This residual equation has the same form as the corresponding matrix result from [11], but holds in this more general setting. It appears to be better suited to dealing with eigenvalue multiplicity than are other approaches. We also argue that when dealing with the lower part of the spectrum of the positive definite operator our choice of the regularization is optimal.

Specifically, we follow the approach of [7, 10, 11] and reverse the trend to show that these finite dimensional results have a lot to offer in the original setting of [21], and in particular as tools for a numerical study of singularly perturbed (integro) differential operators. This paper will heavily use the general construction from [15, 17]. In short we propose, assuming we are given a positive definite and self-adjoint operator  $\mathbf{H}$  and the orthogonal projection  $P$ :

- Construct the positive definite operator  $\mathbf{H}'$  from [15, 17].  $\mathbf{H}'$  is the block diagonal part of  $\mathbf{H}$  with respect to  $P$  and it is given formally as  $\mathbf{H}' = \mathbf{P}\mathbf{H}\mathbf{P} + (\mathbf{I} - \mathbf{P})\mathbf{H}(\mathbf{I} - \mathbf{P})$ . In this paper we shall use the notation  $\mathbf{H}_P = \mathbf{H}'$  to emphasize the dependence on the projection  $P$ .
- Scale the perturbation  $\mathbf{H} - \mathbf{H}_P$  with  $\mathbf{H}_P^{-1/2}$  to obtain the bounded operator  $\delta H_s^P := \mathbf{H}_P^{-1/2}(\mathbf{H} - \mathbf{H}_P)\mathbf{H}_P^{-1/2}$ . This is equivalent to working with the perturbation  $\mathbf{H} - \mathbf{H}_P$  in the dual energy space which is associated to  $\mathbf{H}_P$ .
- Apply an adapted result from [7, 10, 11] to obtain the desired eigenvalue/ vector estimate in the quadratic form setting.

At the end, we would like to emphasize that the original matrix inequalities from [11, 12] have been extensively tested in the process of developing new finite precision eigenvalue software. In the course of this testing a large body of test examples has been generated by judicious random searches as well as by a modification of the known examples from science and engineering. The inequalities have been found to be numerically sharp, as is reported in [12], on numerous test matrices.

**1.2. Relationship to Temple-Kato inequality.** The central theme of this paper is the issue of how to regularize an unbounded eigenvalue problem to obtain an object which can then be algebraically studied. As an introduction to the issue we shortly review other approaches with special emphasis on the regularity issues. This section is meant to provide the motivation for this study and it extends the introduction.

The history of *a posteriori* eigenvalue approximation estimates goes back to [20, 32]. Such inequalities (most recently studied in [26]) have a general form of

$$(1.1) \quad |\text{ERROR}| \leq \text{CONDITIONING} \times \|\text{RESIDUAL}\|^2.$$

Our estimates will have the form of

$$(1.2) \quad \|\text{RESIDUAL}\|_{\text{rel}}^2 \leq |\text{rel}(\text{ERROR})| \leq \text{rel}(\text{CONDITIONING}) \times \|\text{RESIDUAL}\|_{\text{rel}}^2,$$

where the measures  $\|\cdot\|_{\text{rel}}$  and  $|\text{rel}(\text{ERROR})|$  and  $\text{rel}(\text{conditioning})$  denote the appropriate ingredients from matrix (relative) perturbation theory as given in [24].

Let us now be more precise. We shall always work in the background Hilbert space  $\mathcal{H}$ , which is equipped with the scalar product  $(\cdot, \cdot)$  and the norm  $\|\cdot\| = (\cdot, \cdot)^{1/2}$ . Let  $\mathbf{H}$  be a self-adjoint operator which is bounded from below and let  $\psi$  be some vector of the norm one in its *domain of definition*  $\mathcal{D}(\mathbf{H})$ . Define the *Rayleigh-quotient*  $\mu := (\psi, \mathbf{H}\psi)$  and assume<sup>1</sup> that  $\lambda_2(\mathbf{H}) > \mu$  then classical Temple-Kato inequality from [28, Theorem VIII.5, Volume IV pp. 84] reads

$$(1.3) \quad \mu - \frac{(\mathbf{H}\psi, \mathbf{H}\psi) - (\psi, \mathbf{H}\psi)^2}{\lambda_2(\mathbf{H}) - \mu} \leq \lambda_1(\mathbf{H}) \leq \mu.$$

The vector  $r = \mathbf{H}\psi - \mu\psi$  is called the residual (associated to  $\psi$ ) and it holds  $(\mathbf{H}\psi, \mathbf{H}\psi) - (\psi, \mathbf{H}\psi)^2 = \|r\|^2$ . Now it is easy to see that (1.3) implies

$$(1.4) \quad \mu - \lambda_1 \leq \frac{\|\mathbf{H}\psi - \mu\psi\|^2}{\lambda_2 - \mu},$$

---

<sup>1</sup>We are counting the eigenvalues, which are below the infimum of the essential spectrum, in the ascending order according to multiplicity.

which has a general form of (1.1). Here we have used  $\lambda_i = \lambda_i(\mathbf{H})$ ,  $i = 1, 2$  to simplify the notation. The norm  $\|r\|$  can be seen as an “approximation defect” of the vector  $\psi$ , since  $\psi$  is an eigenvector if and only if  $\|r\| = 0$ .

As already stated our estimates have a similar general form (see (1.2)) to the Temple–Kato inequality but are obtained under the assumptions of the perturbation theory for symmetric forms from [21, Chapters VI–VIII]. A consequence of this is that, in a case of a positive definite operator  $\mathbf{H}$ , we are able to directly work with test vectors  $\psi$  from the domain of the symmetric form which is according to [21, Theorem VI-2.23, pp. 331] equal to  $\mathcal{D}(\mathbf{H}^{1/2})$ . Our version of (1.3) also assumes  $\mu < \lambda_2(\mathbf{H})$  but allows  $\psi \in \mathcal{D}(\mathbf{H}^{1/2})$ ,  $\|\psi\| = 1$  and establishes (Theorem 3.4) the estimate<sup>2</sup>

$$(1.5) \quad \frac{\|\mathbf{H}_\mu^{\frac{1}{2}}\psi - \psi\|_{\mathbf{H}^{-1}}^2}{\|\psi\|_{\mathbf{H}^{-1}}^2} \leq \frac{\mu - \lambda_1}{\mu} \leq \frac{\lambda_2 + \mu}{\lambda_2 - \mu} \frac{\|\mathbf{H}_\mu^{\frac{1}{2}}\psi - \psi\|_{\mathbf{H}^{-1}}^2}{\|\psi\|_{\mathbf{H}^{-1}}^2},$$

where  $\|\cdot\|_{\mathbf{H}^{-1}} := \|\mathbf{H}^{-1/2} \cdot\|$  is the classical  $\mathbf{H}^{-1}$ -norm. To recognize the importance of the original Temple–Kato approach, as well as in line with the terminology from Numerical Linear Algebra, see [25, pp. 271], we call all the inequalities which have the form of (1.1) or (1.2) *Temple–Kato like inequalities*. By Temple–Kato approach we mean the notion that high performance eigenvalue estimates should be obtained as a mixture of the *a posteriori* computable measure of the approximation defect  $\|r\|^2$  and the *a priori* assumed quantitative information  $1/(\lambda_2(\mathbf{H}) - \mu) = \max\{|\lambda - \mu|^{-1} : \lambda \in \Sigma(\mathbf{H}) \setminus \{\lambda_1(\mathbf{H})\}\}$  on the conditioning of  $\lambda_1(\mathbf{H})$ .

Let us now discuss (1.5). The measure of `rel(CONDITIONING)` is in this context the so called *relative gap*  $(\lambda_2 - \mu)/(\lambda_2 + \mu)$ , which distinguishes close eigenvalues better than does the *absolute gap*  $(\lambda_2 - \mu)$  from (1.3). Furthermore, both the residual measure  $\|\mathbf{H}_\mu^{\frac{1}{2}}\psi - \psi\|_{\mathbf{H}^{-1}}^2/\|\psi\|_{\mathbf{H}^{-1}}^2$  as well as the relative gap are robust with regard to scaling (e.g. “dimensionless quantities”). Thus, the most important message of (1.5) is the same as in [12, Example 2.1]: The approximation  $\mu$  has completely resolved the eigenvalue  $\lambda_1$  when the (relative) residual measure drops below the relative gap. We also note that (1.5), unlike (1.3) cannot give negative lower bounds to eigenvalues of positive definite operators. On the other hand, the  $\mathbf{H}^{-1}$ -norm is more difficult to evaluate than are the ingredients of (1.5). For a possibility to do this see Remark 3.9, Section 4 and [3, 8, 16]. Approximations to  $\mathbf{H}^{-1}$  norm of the residual can also be computed in a more accessible scalar product, see [10, Remark 7]. Note that the restriction  $\psi \in \mathcal{D}(\mathbf{H})$  from (1.3) excludes—without prior regularization of the problem—the case when  $\psi$  is a continuous piecewise linear function and  $\mathbf{H} = -\Delta$  is the negative *Laplace operator*. On the other hand, in this case is our theory directly applicable and working estimates are explicitly given, accompanied with an argument for their optimality. Furthermore, as an illustration of our matrix theoretic approach to unbounded operators, we will show (in Section 2) a “matrix analytic” way to obtain a variant of the original Temple–Kato inequality.

We close this section by a simple and small numerical example which should illustrate the dichotomy between the (easy) computability and scaling robustness of eigenvalue estimates. We will be comparing the first order estimates (in the approximation defect) from [15, 17] with the second order estimate (1.4). As a model we consider the asymptotic behavior of the family

<sup>2</sup>The conditioning constant  $\frac{\lambda_2 + \mu}{\lambda_2 - \mu}$  is a deliberate overestimate of the optimal constant  $g_1$  from Theorem 3.4. It is a classical ingredient of the *relative perturbation theory*, see [12, 24].

of positive-definite matrices

$$(1.6) \quad \mathbf{H}_\kappa = \begin{bmatrix} \frac{1}{101} & 0 & -\frac{1}{101} \\ 0 & \frac{1}{100} & 0 \\ -\frac{1}{101} & 0 & 1 + \kappa^2 \end{bmatrix}, \quad \kappa \rightarrow \infty.$$

The results of [15, 17], which in this specialization to the matrix  $\mathbf{H}_\kappa$  can be obtained using [11, Theorem 1.1] and a direct computation, use the relative residual measure

$$\eta(\psi) := \frac{\|\mathbf{H}_\mu^{\frac{1}{2}}\psi - \psi\|_{\mathbf{H}^{-1}}}{\|\psi\|_{\mathbf{H}^{-1}}}, \quad \mu = (\psi, \mathbf{H}\psi).$$

Before we proceed note the following geometrical facts. It holds that  $\eta(\psi)$  is equal to the sine of the angle  $\angle(\mathbf{H}\psi, \psi)$  in the scalar product  $(\cdot, \mathbf{H}^{-1}\cdot)$ . We denote this by writing  $\eta(\psi) = \sin \angle(\mathbf{H}\psi, \psi)_{\mathbf{H}^{-1}}$ . It can also be shown that in the scalar product of the background space  $\mathcal{H}$  the identity  $\eta(\psi) = \sin \angle(\mathbf{H}^{1/2}\psi, \mathbf{H}^{-1/2}\psi)$  holds. Furthermore,  $\eta(\psi)$  is also an approximation defect measure, since  $\eta(\psi) \geq 0$  and  $\psi$  is an eigenvector of  $\mathbf{H}$  if and only if  $\eta(\psi) = 0$ .

Now, [15, Theorem 5.1] states that if for  $\psi \in \mathcal{D}(\mathbf{H}^{1/2})$  the assumption  $\eta(\psi) < \frac{\lambda_2(\mathbf{H}) - \mu}{\lambda_2(\mathbf{H}) + \mu}$  holds then

$$(1.7) \quad \frac{|\lambda_1(\mathbf{H}) - \mu|}{\mu} \leq \eta(\psi).$$

This estimate is equivalent with

$$(1.8) \quad (1 - \eta(\psi))\mu \leq \lambda_1(\mathbf{H}) \leq (1 + \eta(\psi))\mu.$$

and since both  $\mu$  and  $\eta(\psi)$  are computable relations (1.7) and (1.8) give both a *lower* as well as an *upper* estimate for  $\lambda_1(\mathbf{H})$ .

Now, take  $\psi = [1 \ 0 \ 0]^*$  as the test vector and compare  $\|r_\kappa\|$ ,  $r_\kappa = \mathbf{H}_\kappa\psi - \mu\psi$  and  $\eta_\kappa(\psi)$ ,  $\eta_\kappa(\psi) := \|\mathbf{H}_\kappa^{\frac{1}{2}}\psi - \psi\|_{\mathbf{H}_\kappa^{-1}}/\|\psi\|_{\mathbf{H}_\kappa^{-1}}$ .

One computes  $\|r_\kappa\| = \frac{1}{101}$  whereas  $\eta_\kappa(\psi) = \frac{1}{\kappa} \frac{\sqrt{2}}{\sqrt{101\kappa^{-2} + 100}}$ . This shows that the second order estimate from (1.4), as opposed to the first order estimate (1.7), does not detect that  $(\mu - \lambda_1(\mathbf{H}_\kappa))/\mu = \frac{1}{101\kappa^2} + O(\frac{1}{\kappa^4}) \rightarrow 0$  as  $\kappa \rightarrow \infty$ . For more details on a numerical comparison of (1.3)–(1.4) and (1.7)–(1.8) see [15, Table 1.1].

Caution has to be exercised when comparing absolute and relative estimates on this example. It is known that absolute and relative estimation theory can (sometimes) yield equivalent estimates, cf. [19]. For instance, in the case of a single lowermost eigenvalue  $\lambda_1(\mathbf{H})$  an inequality which has a similar form as the righthand side inequality from (1.5) can be obtained if one applies (1.4) to the operator  $\mathbf{H}^{-1}$  in the Hilbert space with the  $\mathbf{H}$ -scalar product  $(\cdot, \cdot)_\mathbf{H} := (\mathbf{H}^{1/2}\cdot, \mathbf{H}^{1/2}\cdot)$ . However, such approaches which first derive the eigenvalue estimates in the background (absolute) scalar product and then scale the operator at hand to fit this framework do not provide a proof of the optimality of the estimates for parameter dependent problems. Furthermore, our approach is more natural for the treatment of the eigenvalue multiplicity which can be seen on the new block-operator residual equation which yields error estimates that utilize any unitary invariant norm of the block operator residual. Such *a posteriori* estimates and block operator residual equations did not appear before in the context of the eigenvalue estimation for unbounded operators. We also note that the numerical examples, reported in [12], indicate

that our choice of the (relative) residual measure  $\eta(\psi)$ , as well as the choice of the measure of the conditioning (relative gap) yield numerically sharp eigenvalue estimates. Bridging the relative perturbation theory from [7, 10, 11] with the theory of eigenvalue estimation for unbounded operators is the declared aim of this work. In addition to that we will outline a new general framework for analyzing asymptotic exactness of eigenvalue estimators for parameter dependent eigenvalue problems. On our simple  $3 \times 3$  example this general result reads

$$\lim_{\kappa \rightarrow \infty} \frac{\frac{\mu - \lambda_1(\mathbf{H}_\kappa)}{\mu}}{\eta_\kappa^2(\psi)} = 1$$

and the rate of the convergence appears to be rather rapid. We will also show, by comparing the block operator residual equation which yields (1.4) with the residual equation that yields (1.5), that the approach of the relative perturbation theory—e.g. first scale and then estimate rather than as in the absolute approach where one first estimates and then scales—is the right one when estimating the lower part of the spectrum of a positive definite unbounded operator.

## 2. A PERTURBATION APPROACH TO RAYLEIGH–RITZ ESTIMATES

We follow the general notational conventions and the terminology of [21, Chapters VI–VIII]. Minor differences are contained in the following list of notation and terminology.

- $\mathcal{H}$  ... is an infinite dimensional Hilbert space, can be both real or complex
- $(\cdot, \cdot); \|\cdot\|$  ... the scalar product on  $\mathcal{H}$ , linear in the second argument and anti-linear (when  $\mathcal{H}$  is complex) in the first; the norm on  $\mathcal{H}$
- $\mathcal{H}_1 \oplus \mathcal{H}_2$  ... the direct sum of the Hilbert spaces  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , for any  $x \in \mathcal{H}_1 \oplus \mathcal{H}_2$  we have  $x = x_1 \oplus x_2 = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  for  $x_i \in \mathcal{H}_i, i = 1, 2$
- $\Sigma(\mathbf{H}), \Sigma_{ess}(\mathbf{H}); \lambda_{ess}(\mathbf{H})$  ... the spectrum and the essential spectrum of  $\mathbf{H}$ ; the infimum of the essential spectrum of  $\mathbf{H}$
- $A \leq B$  ... order relation between self-adjoint operators (matrices), is equivalent with the statement that  $B - A$  is positive
- $\mathcal{L}(\mathcal{H}); \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$  ... the space of bounded linear operators on  $\mathcal{H}$ , which is equipped with the norm  $\|\cdot\|$ ; the space of bounded linear operators from  $\mathcal{H}_1$  to  $\mathcal{H}_2$
- $\mathbf{R}(X), \mathbf{N}(X)$  ... the range and the null space of the linear operator  $X$
- $P, P_\perp$  ... the orthogonal projections  $P$  and  $P_\perp := \mathbf{I} - P$
- $j_{()} \dots$  a permutation of  $\mathbb{N}$
- $\text{diag}(M, W)$  ... the block diagonal operator matrix with the operators  $M, W$  on its diagonal. The operators  $M, W$  can be both bounded and unbounded. The same notation is used to define the diagonal  $m \times m$  matrix  $\text{diag}(\alpha_1, \dots, \alpha_m)$ , with  $\alpha_1, \dots, \alpha_m$  on its diagonal.
- $s_1(A) \geq s_2(A) \geq \dots, s_{\max(A)}, s_{\min(A)}$  ... the singular values of the compact operator  $A$  ordered in the descending order according to multiplicity, the minimal (if it exists) and the maximal singular value of  $A$
- $\|\cdot\|$  ... a unitary invariant or operator cross norm of the operator  $X$ . Since  $\|\cdot\|$  depends only on the singular values of the operator, we do not notationally distinguish between the instances of the norm  $\|\cdot\|$  on  $\mathcal{L}(\mathcal{H}), \mathcal{L}(\mathbf{R}(P)), \mathcal{L}(\mathbf{R}(P))^\perp$ , or such.

Precise properties of a unitary invariant norm will be listed in Section 3, for further details see [30].

- $\text{tr}(X)$ ,  $\|X\|_{HS}$  ... the trace a the Hilbert–Schmidt norm of the operator  $X$ , it holds  $\|X\|_{HS} = \sqrt{\text{tr}(X^*X)}$ , see [30]

As a general policy to simplify the notation we shall always drop indices when there is no danger of confusion.

We will generically assume that we have a closed, symmetric and semibounded from below form  $h$  with the dense domain  $\mathcal{Q}(h) \subset \mathcal{H}$  as given in [21, (VI.1.5)–(VI.1.11), pp. 308–310]. The form  $h$  which has a strictly positive lower bound will be called *positive-definite*. This is also a small departure from the terminology of [21, Section VI.2, pp. 310]. Such form  $h$  defines the self-adjoint and positive definite operator  $\mathbf{H}$  in the sense of [21, Theorem VI.2.23, pp. 331]. Furthermore, the operator  $\mathbf{H}$  is densely defined with the domain  $\mathcal{D}(\mathbf{H}) \subset \mathcal{Q}(h)$  and  $\mathcal{D}(\mathbf{H}^{1/2}) = \mathcal{Q}(h)$ . We also generically assume that  $\mathbf{H}$  has discrete eigenvalues  $\lambda_1(\mathbf{H}) \leq \dots \leq \lambda_m(\mathbf{H}) \leq \dots < \lambda_{\text{ess}}(\mathbf{H})$ , where we count the eigenvalues according to multiplicity. Another departure from the terminology of Kato is that we use  $h(\psi, \phi)$  to denote the value of  $h$  on  $\psi, \phi \in \mathcal{Q}(h)$ , but we write  $h[\psi] := h(\psi, \psi)$  for the associated *quadratic form*  $h[\cdot]$ . We also emphasize that we use  $\cdot^*$  to denote the adjoint both in the real as well as in the complex Hilbert space  $\mathcal{H}$  as is customary in [21, Chapters VI–VIII].

Let us now fix our Rayleigh–Ritz terminology and outline the main construction from [15, 17]. We assume that we are interested in approximating the eigenvalue  $\lambda(\mathbf{H})$  of finite multiplicity  $m \in \mathbb{N}$ . Instead of only one test vector, as was the situation in (1.7), we now need a *test subspace* of dimension  $m$ .

Let therefore  $P$  be an orthogonal projection such that  $\dim \mathcal{R}(P) = m$  and  $\mathcal{R}(P) \subset \mathcal{Q}(h)$ . We call  $\mathcal{R}(P)$  the test subspace for (the approximation of)  $\lambda(\mathbf{H})$ . The operator  $\Xi \in \mathcal{L}(\mathcal{R}(P))$ ,  $\Xi = (\mathbf{H}^{1/2}P|_{\mathcal{R}(P)})^* \mathbf{H}^{1/2}P|_{\mathcal{R}(P)}$  will be called the (generalized) *Rayleigh quotient*. Its eigenvalues  $\mu_1 \leq \dots \leq \mu_m$  will be called the *Ritz values* from the *test subspace*  $\mathcal{R}(P)$  and the vectors  $u_i \in \mathcal{R}(P)$ ,  $\Xi u_i = \mu_i u_i$ ,  $\|u_i\| = 1$  will be called the *Ritz vectors*. We also define the operator  $\mathbf{W} : \mathcal{R}(P)^\perp \rightarrow \mathcal{R}(P)^\perp$  as the one which is defined in  $\mathcal{R}(P)^\perp$  by the form  $h(P_\perp \cdot, P_\perp \cdot)$  in the sense of [21, Theorem VI-2.23, pp. 331].

**2.1. A variant of the Temple–Kato inequality.** Before we outline the main form theoretic construction from [15, 17] let us illustrate our “matrix theoretic” approach to spectral theory by proving a variant of (1.1) in a case when  $\lambda_1(\mathbf{H})$  has a finite multiplicity  $m$ .

Let  $\mathbf{H}$  be a self-adjoint operator which is bounded from below. Let further, counting the eigenvalues according to multiplicity,  $\lambda := \lambda_1(\mathbf{H}) = \lambda_m(\mathbf{H}) < \lambda_{m+1}(\mathbf{H})$ ,  $\lambda \neq 0$  and we assume that we have a test subspace  $\mathcal{R}(P) \subset \mathcal{D}(\mathbf{H})$ ,  $\dim \mathcal{R}(P) = m$ . The environment space  $\mathcal{H}$  can be decomposed as  $\mathcal{H} = \mathcal{R}(P) \oplus \mathcal{R}(P)^\perp$  and  $\mathbf{H}$  can be represented as a block-operator matrix<sup>3</sup>

$$(2.1) \quad \mathbf{H} = \begin{bmatrix} \Xi & K^* \\ K & \mathbf{W} \end{bmatrix},$$

---

<sup>3</sup>For more on block operator matrices see for instance [6].

where  $K \in \mathcal{L}(\mathbf{R}(P), \mathbf{R}(P)^\perp)$ ,  $K = P_\perp \mathbf{H} P|_{\mathbf{R}(P)}$ . Using the standard result<sup>4</sup> [6, Theorem 5.1] one obtains that there exist  $m$ -eigenvalues  $\lambda_{j_1}(\mathbf{H}) \leq \dots \leq \lambda_{j_m}(\mathbf{H})$  such that

$$|\lambda_{j_i}(\mathbf{H}) - \mu_i| \leq \|K\|, \quad i = 1, \dots, m.$$

If we further assume that  $\|K\| < \lambda_{m+1}(\mathbf{H}) - \mu_m$ , then

$$|\lambda_i(\mathbf{H}) - \mu_i| \leq \|K\|, \quad i = 1, \dots, m.$$

follows from the perturbation construction of [6, Theorem 5.1 and Remark 5.2]. Furthermore, the spectral calculus for the self-adjoint operator  $\mathbf{W}$  and the min-max formulae yield

$$\|(\mathbf{W} - \lambda \mathbf{I})^{-1}\| \leq \frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m - \|K\|}.$$

The assumption  $\mathbf{R}(P) \subset \mathcal{D}(\mathbf{H})$  allows us to justify the following matrix representation

$$\mathbf{H} - \lambda \mathbf{I} = \begin{bmatrix} \mathbf{I} & K^* \mathbf{B}_{\text{abs}}^{-1} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \Xi - \lambda \mathbf{I} - K^* \mathbf{B}_{\text{abs}}^{-1} K & \mathbf{0} \\ \mathbf{B}_{\text{abs}} & \mathbf{B}_{\text{abs}}^{-1} K \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{B}_{\text{abs}}^{-1} K & \mathbf{I} \end{bmatrix}, \quad \mathbf{B}_{\text{abs}} := (\mathbf{W} - \lambda \mathbf{I}).$$

We can now use a generalization of the so called Wilkinson's trick from [27, pp. 183], which will be stated explicitly as Theorem 3.1 below, to conclude that

$$m = \dim \mathbf{N}(\mathbf{H} - \lambda \mathbf{I}) = \dim \mathbf{N}(\text{diag}(\Xi - \lambda \mathbf{I} - K^* \mathbf{B}_{\text{abs}}^{-1} K, \mathbf{B}_{\text{abs}}))$$

and this can only happen if

$$(2.2) \quad \Xi - \lambda \mathbf{I} = K^* (\mathbf{W} - \lambda \mathbf{I})^{-1} K.$$

Furthermore, we establish

$$(2.3) \quad \|\text{diag}(\mu_i - \lambda)\| \leq \frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m - \|K\|} \|K\| \|K\|,$$

where  $\|\cdot\|$  is any unitary invariant norm. A similar inequality holds for discrete eigenvalues which are located in the interior of  $\Sigma(\mathbf{H})$ . In particular, if we denote the (single) Ritz vector residuals by  $r_i := \mathbf{H}u_i - \mu_i u_i$  and apply the trace operator  $\text{tr}(\cdot)$  on (2.2), we obtain<sup>5</sup>

$$(2.4) \quad \sum_{i=1}^m |\mu_i - \lambda| \leq \frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m - \|K\|} \sum_{i=1}^m \|r_i\|^2.$$

This generalizes the estimate from (1.3) to the case in which  $\lambda_1(\mathbf{H})$  has the multiplicity  $m$ . The quotient  $\frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m - \|K\|}$  is numerically inferior to  $\frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m}$ , but (2.3) holds for any unitary invariant norm. This is a mechanism which allows us to individually treat Ritz vectors of different approximation properties. Furthermore, in view of the discussion from [5, pp. 8] estimates (2.3) give significant new information when compared just with (2.4). It could be argued that  $\frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m - \|K\|}$  is a reasonable ingredient of the estimates since, in typical situations, one uses

<sup>4</sup>Also known as Kahan's residual theorem in the case of the test subspace  $\mathbf{R}(P)$  of Krylov-Weinstein inequality in the case of one test vector  $\psi$ , see [4, 27].

<sup>5</sup>The inequality (2.4) appeared with a better gap estimate in the bounded operator setting in [26], whereas the inequality (2.3) appeared in the matrix setting in [31]. The significance of this inequalities in this paper is to introduce the Schur complement technique which will be the main tool later.

[6, Theorem 5.1 and Remark 5.2] to obtain a bound for  $\frac{1}{\lambda_{m+1}(\mathbf{H}) - \mu_m}$ . Also, see the discussion on [4, pp. 305]. For a way to compute  $\|K\|$  see [5, Section 9].

**2.2. The symmetric form approach.** The previous computation can not be justified in the case in which  $\mathbf{R}(P) \subset \mathcal{Q}(h) = \mathcal{D}(\mathbf{H}^{1/2})$  but  $\mathbf{R}(P) \not\subset \mathcal{D}(\mathbf{H})$ . Precisely this is the case in which we are interested.

Let us now outline the main perturbation construction from [15, 17]. We start by defining the positive definite form

$$h_P(u, v) = h(Pu, Pv) + h(P_\perp u, P_\perp v), \quad u, v \in \mathcal{Q}(h)$$

and the self-adjoint operator  $\mathbf{H}_P$  which is defined by  $h_P$  in the sense of [21, Theorem VI-2.23, pp. 331]. The operator  $\mathbf{H}_P$  and the form  $h_P$  are called the *P-diagonal* part of  $\mathbf{H}$  and  $h$ , respectively. We also define the form

$$\delta h^P(u, v) = h(Pu, P_\perp v) + h(P_\perp u, Pv), \quad u, v \in \mathcal{Q}(h),$$

which is an approximation defect in  $\mathbf{R}(P)$ , since  $\mathbf{R}(P)$  is an invariant subspace of  $\mathbf{H}$  if and only if  $\delta h^P \equiv 0$ , for a proof see [17]. Furthermore, it was shown in [15, 17] that

- (1)  $\mathbf{R}(P)$  reduces  $\mathbf{H}_P$
- (2)  $\Xi = P\mathbf{H}_P P|_{\mathbf{R}(P)}$
- (3)  $\mathbf{R}(\mathbf{H}^{-1} - \mathbf{H}_P^{-1})$  is finite dimensional which implies that  $\Sigma_{ess}(\mathbf{H}) = \Sigma_{ess}(\mathbf{H}_P)$  according to the Weyl theorem.

The properties 1), 2) and 3) imply that  $\mu_i \in \Sigma(\mathbf{H}_P)$  and  $\lambda_{ess}(\mathbf{H}) = \lambda_{ess}(\mathbf{H}_P)$  together with the assumption  $\mu_i < \lambda_{ess}(\mathbf{H})$  yields that  $\mu_i$  are the eigenvalues of the operator  $\mathbf{H}_P$  with finite multiplicity. We are setting the scene for an application of the relative perturbation theory from [21, Chapters VI–VIII] and so we will be able, regardless of the fact that<sup>6</sup>  $\mathcal{D}(\mathbf{H}) \neq \mathcal{D}(\mathbf{H}_P)$ , to interpret  $\mathbf{H}$  as a perturbation of  $\mathbf{H}_P$  and thus bring  $\mu_i$  in connection with some component of  $\Sigma(\mathbf{H})$ .

This was the main line of argument in [15, 17]. Although some of the technical results about  $\mathbf{H}_P$ , which we shall now state are not explicitly given in [15, 17] we present them here without proof. However, all of their proofs are obtainable as minor modifications of the arguments from [15, 17] and do not bring any new information.

Let us now look into the structure of this construction in more detail. According to [15, Theorem 4.5] the form  $\delta h_s^P(\cdot, \cdot) := \delta h^P(\mathbf{H}_P^{-1/2} \cdot, \mathbf{H}_P^{-1/2} \cdot)$  defines the bounded operator  $\delta H_s^P$  and

$$(2.5) \quad \|\delta H_s^P\| = \max_{\psi \in \mathbf{R}(P)} \frac{(\psi, \mathbf{H}^{-1}\psi) - (\psi, \mathbf{H}_P^{-1}\psi)}{(\psi, \mathbf{H}^{-1}\psi)}.$$

To examine  $\delta H_s^P$  in further detail define

$$(2.6) \quad \eta_i(P) := \left[ \max_{\substack{\mathcal{S} \subset \mathbf{R}(P), \\ \dim(\mathcal{S})=m-i+1}} \min \left\{ \frac{(\psi, \mathbf{H}^{-1}\psi) - (\psi, \mathbf{H}_P^{-1}\psi)}{(\psi, \mathbf{H}^{-1}\psi)} \mid \psi \in \mathcal{S}, \|\psi\| = 1 \right\} \right]^{1/2},$$

<sup>6</sup>In fact, it is even possible that  $\mathcal{D}(\mathbf{H}) \cap \mathcal{D}(\mathbf{H}_P) = \{0\}$  and the form approach is still applicable.

for  $i = 1, \dots, m$ . Obviously,  $\|\delta H_s^P\| = \eta_m(P)$ . A more detailed assessment of the proof of [15, Theorem 4.1] yields the following lemma.

**Lemma 2.1.** *Let  $h$  be positive definite, and let  $\mathcal{R}(P)$  be the test subspace such that  $\dim \mathcal{R}(P) = m$ . Assume further that*

$$\eta_{r+1}(P) \leq \eta_{r+2}(P) \leq \dots \leq \eta_m(P)$$

*are all nonzero  $\eta_i(P)$  from (2.6). Then  $\eta_m(P) < 1$  and  $\pm \eta_{r+1}(P), \dots, \pm \eta_m(P)$  are all non-zero eigenvalues of  $\delta H_s^P$ . Furthermore,  $\eta_{r+1}(P), \dots, \eta_m(P)$  are all non-zero singular values of the operator  $K_s = \delta H_s^P P|_{\mathcal{R}(P)} \in \mathcal{L}(\mathcal{R}(P), \mathcal{R}(P)^\perp)$ .*

*Proof.* The proof of this lemma is implicitly contained in the proof of [15, Theorem 4.1]. We leave out most of the technical details. We only explicitly present arguments that  $\eta_m(P) < 1$ . This fact was first established, in the matrix case, by Z. Drmač, Zagreb. Since  $\mathcal{R}(P)$  reduces  $\mathbf{H}_P$  we have

$$h(\mathbf{H}_P^{-1}f, v) = h_P(\mathbf{H}_P^{-1}f, v) = (f, v), \quad v \in \mathcal{R}(P),$$

i.e.  $\mathbf{H}_P^{-1}f$  is a Galerkin approximation from the subspace  $\mathcal{R}(P)$  to  $\psi = \mathbf{H}^{-1}f$ , which solves the problem  $\mathbf{H}\psi = f$ . With this in hand one computes

$$(2.7) \quad h(\mathbf{H}^{-1}f - \mathbf{H}_P^{-1}f, \mathbf{H}^{-1}g - \mathbf{H}_P^{-1}g) = (f, \mathbf{H}^{-1}g) - (f, \mathbf{H}_P^{-1}g), \quad f, g \in \mathcal{R}(P).$$

This implies  $(f, \mathbf{H}^{-1}f) \geq (f, \mathbf{H}_P^{-1}f) \geq 0$ ,  $f \in \mathcal{R}(P)$  and  $0 \leq \eta_i(P) \leq 1$  follows. Assume  $\eta_m(P) = 1$ , then there exists  $f \in \mathcal{R}(P) \setminus \{0\}$  such that  $(f, \mathbf{H}_P^{-1}f) = 0$ . This is an obvious contradiction with the fact that  $\mathbf{H}_P$  is positive definite.  $\square$

Now, set formally  $\lambda_0(\mathbf{H}) := 0$ ,  $\mathfrak{g}_0 := \infty$  and define

$$(2.8) \quad \mathfrak{g}_q := \min \left\{ |\lambda_q(\mathbf{H}) - \mu| \mu^{-1} : \mu \in \Sigma(\mathbf{H}_P) \setminus \{\mu_1, \dots, \mu_m\} \right\}$$

$$(2.9) \quad \gamma_s(\lambda_q) := \min \left\{ \frac{\lambda_{q+m}(\mathbf{H}) - \mu_m}{\lambda_{q+m}(\mathbf{H}) + \mu_m}, \frac{\mu_1 - \lambda_{q-1}(\mathbf{H})}{\mu_1 + \lambda_{q-1}(\mathbf{H})} \right\},$$

for  $q \in \mathbb{N}$ . These quantities—which measure the sensitivity of the eigenvalue  $\lambda_q(\mathbf{H})$ —will play a role in the statement of the theorems in the next section. In the rest of the section we suppress the dependence of quantities on  $\mathbf{H}$  and  $P$  in the notation.

We now relate  $\gamma_s(\lambda_q)$  and  $\mathfrak{g}_q$  to  $\eta_i$ . The main result of [15] established that given  $\mathcal{R}(P) \subset \mathcal{Q}(h)$ ,  $\dim \mathcal{R}(P) = m$  and  $\mu_i < \lambda_{\text{ess}}$  there exist  $m$  eigenvalues  $\lambda_{j_1} \leq \dots \leq \lambda_{j_m}$  such that

$$\frac{|\lambda_{j_i} - \mu_i|}{\mu_i} \leq \eta_m, \quad i = 1, \dots, m$$

holds. Under an additional assumption on the location of the unwanted component of the spectrum we can localize the approximated eigenvalues and obtain that if e.g.  $\eta_m(1 - \eta_m)^{-1} < (\lambda_{m+1} - \mu_m)(\lambda_{m+1} + \mu_m)^{-1}$  then

$$(2.10) \quad \frac{|\lambda_i - \mu_i|}{\mu_i} \leq \eta_m, \quad i = 1, \dots, m.$$

This assumption is similar to the assumption of the Temple–Kato inequality (1.3). For higher eigenvalues we have the following variant of [15, Theorems 5.1 and 5.2], which we present without proof.

**Lemma 2.2.** *Let  $\mathbf{R}(P)$  be the test subspace for the positive definite form  $h$  and let  $\dim \mathbf{R}(P) = m$  and  $q \in \mathbb{N}$ . If  $\frac{\eta_m}{1-\eta_m} < \gamma_s(\lambda_q)$  then  $\mathbf{g}_q > 0$  and in particular<sup>7</sup>*

$$\mathbf{g}_q \geq \min \left\{ \frac{\mu_1(1-\eta_m) - (1 + \frac{\eta_m}{1-\eta_m})\lambda_{q-1}}{(1 + \frac{\eta_m}{1-\eta_m})\lambda_{q-1}}, \frac{(1 - \frac{\eta_m}{1-\eta_m})\lambda_{q+m} - (1 + \eta_m)\mu_m}{(1 - \frac{\eta_m}{1-\eta_m})\lambda_{q+m}} \right\}.$$

### 3. TEMPLE-KATO LIKE INEQUALITY IN THE PRESENCE OF RITZ VALUE CLUSTERS

We now present the main contribution of this article. We will derive relative eigenvalue estimates in the presence of Ritz value clusters.

In this section we will need to elaborate on the notion of the unitary invariant operator norm (also known as symmetric or cross operator norms, cf. [21, 30] and the references therein). This will allow us to extract more information from  $\mathbf{R}(P)$  than what is contained in  $\eta_m(P) = \|K_s\|$ . In this section we will be dealing with only one orthogonal projection  $P$ , and so we simply write  $\eta_i := \eta_i(P)$ ,  $i = 1, \dots, m$ , whenever there is no danger of confusion. Furthermore, we write  $\lambda_i := \lambda_i(\mathbf{H})$ ,  $i \in \mathbb{N}$  to simplify the notation.

To say that the norm is *unitary invariant* on  $\mathcal{S} \subset \mathcal{L}(\mathcal{H})$  means that, beside the usual properties of any norm, it additionally satisfies:

**(i):** If  $B \in \mathcal{S}$ ,  $A, C \in \mathcal{L}(\mathcal{H})$  then  $ABC \in \mathcal{S}$  and

$$(3.1) \quad \|\cdot\| \leq \|A\| \|\cdot\| B \|\cdot\| C.$$

**(ii):** If  $A$  has rank 1 then  $\|\cdot\| = \|A\|$ , where  $\|\cdot\|$  always denotes the standard operator norm on  $\mathcal{L}(\mathcal{H})$ .

**(iii):** If  $A \in \mathcal{S}$  and  $U, V$  are unitary on  $\mathcal{H}$ , then  $UAV \in \mathcal{S}$  and

$$(3.2) \quad \|\cdot\| = \|A\|.$$

**(iv):**  $\mathcal{S}$  is complete under the norm  $\|\cdot\|$ .

The subspace  $\mathcal{S}$  is defined as a  $\|\cdot\|$ -closure of the set of all degenerate operators in  $\mathcal{L}(\mathcal{H})$ . Such  $\mathcal{S}$  is an ideal in the algebra  $\mathcal{L}(\mathcal{H})$ , cf. [30].

A typical example of such a norm is the Hilbert–Schmidt norm  $\|\cdot\|_{HS}$ . A bounded operator  $H : \mathcal{H} \rightarrow \mathcal{H}$  is a Hilbert–Schmidt operator if  $H^*H$  is trace class and then, cf. [21, Ch. X.1.3],  $\|H\|_{HS} := \sqrt{\text{tr } H^*H} = [\sum_{i=1}^{\infty} s_i(H)^2]^{1/2}$ .

Before we turn to the main theorem, let us give an alternative—more common—definition for the approximation defects  $\eta_i(P)$ . To this end we further exploit the Galerkin orthogonality property of the  $P$ -diagonal part of  $h$  and in particular the ramifications of relation (2.7) from Lemma 2.1. For any  $f \in \mathbf{R}(P)$  we have

$$h[\mathbf{H}^{-1}f - \mathbf{H}_P^{-1}f] = \left[ \sup_{\phi \in \mathcal{Q}(h) \setminus \{0\}} \frac{|h(\mathbf{H}^{-1}f - \mathbf{H}_P^{-1}f, \phi)|}{h[\phi]^{1/2}} \right]^2 = \|\mathbf{H}\Xi^{-1}f - f\|_{\mathbf{H}^{-1}}^2.$$

With this we can write (2.6) in an alternative form

$$(3.3) \quad \eta_i(P) := \left[ \max_{\substack{\mathcal{S} \subset \mathbf{R}(P), \\ \dim(\mathcal{S})=m-i+1}} \min \left\{ \frac{\|\mathbf{H}\Xi^{-1}\psi - \psi\|_{\mathbf{H}^{-1}}^2}{\|\psi\|_{\mathbf{H}^{-1}}^2} \mid \psi \in \mathcal{S}, \|\psi\| = 1 \right\} \right]^{1/2},$$

<sup>7</sup>Here we assume  $\frac{c}{\infty} = 0$ , for  $c > 0$ .

for  $i = 1, \dots, m$ .

**3.1. Operator matrices and the Wilkinson's trick.** As a first step we shall outline the Wilkinson's trick and state our adaptation of this result as a theorem. This result yielded (2.2). We shall then proceed to prove eigenvalue estimates. Let us now generalize the Wilkinson's trick to operator matrices, cf. [27, p. 183].

**Theorem 3.1** (Wilkinson's trick). *Let  $A : \mathcal{H}_1 \rightarrow \mathcal{H}_1$  and  $X : \mathcal{H}_2 \rightarrow \mathcal{H}_1$  be bounded operators and let  $A$  be self-adjoint. Assume further that  $\mathbf{B} : \mathcal{H}_2 \rightarrow \mathcal{H}_2$  is self-adjoint and that it has a bounded inverse and define  $\mathbf{M} = \begin{bmatrix} A & X \\ X^* & \mathbf{B} \end{bmatrix}$ , to be understood as operator on  $\mathcal{H}_1 \oplus \mathcal{H}_2$ . If  $\dim \mathbf{N}(\mathbf{M}) = \dim \mathcal{H}_1 < \infty$  then*

$$A = X\mathbf{B}^{-1}X^*.$$

*Proof.* We shall adapt the Schur-complement technique from [27, p. 183]. Since  $\mathbf{B}^{-1}$  is assumed to be bounded we can write

$$(3.4) \quad \mathbf{M} = \begin{bmatrix} \mathbf{I} & X\mathbf{B}^{-1} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} A - X\mathbf{B}^{-1}X^* & 0 \\ 0 & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{B}^{-1}X^* & \mathbf{I} \end{bmatrix} = S\mathbf{D}S^*.$$

Both of the operator matrices

$$S = \begin{bmatrix} \mathbf{I} & X\mathbf{B}^{-1} \\ 0 & \mathbf{I} \end{bmatrix}, \quad S^{-1} = \begin{bmatrix} \mathbf{I} & -X\mathbf{B}^{-1} \\ 0 & \mathbf{I} \end{bmatrix}$$

define bounded operators on  $\mathcal{H}_1 \oplus \mathcal{H}_2$ , and so  $\mathbf{D} = S^{-1}\mathbf{M}S^{-*}$ . This implies that  $\mathcal{D}(\mathbf{M}) = \mathcal{D}(\mathbf{D})$  and as a consequence of a simple dimension counting we obtain that

$$(3.5) \quad \dim \mathbf{N}(\mathbf{M}) = \dim \mathbf{N}(\mathbf{D}) < \infty$$

Since  $\mathbf{B}$  has a bounded inverse (3.5) can only be true if  $A - X\mathbf{B}^{-1}X^* = 0$ . This is the so called Wilkinson's trick and it proves the statement of the theorem.  $\square$

**Remark 3.2.** Note that the theorem remains valid if we only assume that  $\mathbf{B}$  is injective and  $\mathbf{B}^{-1}X^*$  is bounded. In this case we conclude that  $A = X(\mathbf{B}^{-1}X^*)$ . In the case when  $\mathcal{H}_1$  is infinite dimensional the dimension counting cannot be used to prove the result. Some spectral properties of Schur complements in a general situation can be found in [22].

**Theorem 3.3.** *Let  $\mathbf{R}(P)$  be the test subspace for the positive definite form  $h$ , as defined in Section 2, and let  $\dim \mathbf{R}(P) = m$  and  $q \in \mathbb{N}$ . Assume that  $\lambda_{q-1} < \lambda_q = \lambda_{q+m-1} < \lambda_{q+m}$  and  $\frac{\eta_m}{1-\eta_m} < \gamma_s(\lambda_q)$  hold then*

$$(3.6) \quad \|\mathbf{I} - \lambda_q \Xi^{-1}\| \leq \frac{\eta_m}{\mathfrak{g}_q} \|\text{diag}(\eta_1, \dots, \eta_m)\|.$$

*In particular, for  $\|\cdot\| = \|\cdot\|_{HS}$  and  $\mu_i \in \Sigma(\Xi)$  we have the estimate*

$$(3.7) \quad \left[ \sum_{i=1}^m \frac{(\lambda_q - \mu_i)^2}{\mu_i^2} \right]^{1/2} \leq \frac{\eta_m}{\mathfrak{g}_q} \left[ \eta_1^2 + \dots + \eta_m^2 \right]^{1/2}.$$

*Proof.* Let the form  $h_P$  be the  $P$ -diagonal part of  $h$ . A modification of [15, Theorems 5.1 and 5.2] implies that

$$h(\mathbf{H}_P^{-1/2} \cdot, \mathbf{H}_P^{-1/2} \cdot) - \lambda_q(\mathbf{H}_P^{-1/2} \cdot, \mathbf{H}_P^{-1/2} \cdot)$$

defines the bounded operator  $H_s(\lambda_q)$ , which allows the operator matrix representation

$$(3.8) \quad H_s(\lambda_q) = \begin{bmatrix} \mathbf{I} - \lambda_q \Xi^{-1} & K_s^* \\ K_s & \mathbf{I} - \lambda_q \mathbf{W}^{-1} \end{bmatrix},$$

with respect to  $\mathcal{H} = \mathcal{R}(P) \oplus \mathcal{R}(P)^\perp$ . Now, Lemma 2.2 implies that  $\mathbf{I} - \lambda_q \mathbf{W}^{-1}$  is invertible and we may use the Wilkinson's trick to derive quadratic estimates (for some further technical details see Lemma 2.2). In particular we have  $\|(\mathbf{I} - \lambda_q \mathbf{W}^{-1})^{-1}\| = \frac{1}{\mathfrak{g}_q} < \infty$ . Now temporarily set  $\mathbf{B}_{\text{rel}} := (\mathbf{I} - \lambda_q \mathbf{W}^{-1})$ , then

$$H_s(\lambda_q) = \begin{bmatrix} \mathbf{I} & K_s^* \mathbf{B}_{\text{rel}}^{-1} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} (\mathbf{I} - \lambda_q \Xi^{-1}) - K_s^* \mathbf{B}_{\text{rel}}^{-1} K_s & 0 \\ 0 & \mathbf{B}_{\text{rel}} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{B}_{\text{rel}}^{-1} K_s & \mathbf{I} \end{bmatrix}$$

and Theorem 3.1 yields

$$(3.9) \quad \mathbf{I} - \lambda_q \Xi^{-1} = K_s^* (\mathbf{I} - \lambda_q \mathbf{W}^{-1})^{-1} K_s.$$

Property (3.1) of a unitary invariant norm  $\|\cdot\|$  implies

$$\|\mathbf{I} - \lambda_q \Xi^{-1}\| \leq \|K_s\| \|(\mathbf{I} - \lambda_q \mathbf{W}^{-1})^{-1} K_s\|.$$

We apply (3.2) and Theorem 2.1 on the last inequality to complete the proof.  $\square$

The estimate of Theorem 3.3 was an equality up to (3.9). So, there is more information in (3.9) than we have used so far.

**Theorem 3.4.** *If  $\mu_m < \lambda_{m+1}$  and  $\lambda_1 = \lambda_m$  then*

$$(3.10) \quad \|\text{diag}(\eta_1^2, \dots, \eta_m^2)\| \leq \|\mathbf{I} - \lambda_q \Xi^{-1}\| \leq \frac{1}{\mathfrak{g}_1} \|\text{diag}(\eta_1^2, \dots, \eta_m^2)\|$$

$$(3.11) \quad \sum_{i=1}^m \eta_i^2 \leq \sum_{i=1}^m \frac{\mu_i - \lambda_i}{\mu_i} \leq \frac{1}{\mathfrak{g}_1} \sum_{i=1}^m \eta_i^2.$$

*Proof.* The assumption  $\mu_m < \lambda_{m+1}$  implies  $\mathfrak{g}_1 > 0$ . Now, this combined with Lemma 2.2 yields

$$(3.12) \quad K_s^* K_s \leq \mathbf{I} - \lambda_q \Xi^{-1} \leq \mathfrak{g}_1^{-1} K_s^* K_s.$$

The conclusions now readily follow by an application of a norm  $\|\cdot\|$  and the trace operator  $\text{tr}(\cdot)$  on (3.12).  $\square$

We now relate  $\mathfrak{g}_1$  to the standard relative gap which was extensively studied in the relative perturbation theory. The result of Lemma 2.2 can be easily improved if we concentrate on the lower part of the spectrum. This is a reasonable assumption.

**Corollary 3.5.** *Assume that  $\lambda_1 = \dots = \lambda_m < \lambda_{m+1}$  and  $\eta_m < (\lambda_{m+1} - \mu_m)(\lambda_{m+1} + \mu_m)^{-1}$  then  $\mathfrak{g}_1 \geq \frac{\lambda_{m+1} - \mu_m}{\lambda_{m+1} + \mu_m}$  and in particular*

$$(3.13) \quad \|\text{diag}(\eta_1^2, \dots, \eta_m^2)\| \leq \|\mathbf{I} - \lambda_q \Xi^{-1}\| \leq \frac{\lambda_{m+1} + \mu_m}{\lambda_{m+1} - \mu_m} \|\text{diag}(\eta_1^2, \dots, \eta_m^2)\|.$$

**Remark 3.6.** Theorem 3.4 illustrates why this “relative” or form approach to the Temple–Kato inequality is more natural than the one which yielded (2.2). The operator  $\mathbf{B}_{\text{rel}}$  in the relative block operator residual equation (3.9) is such that  $\mathbf{I} \leq \mathbf{B}_{\text{rel}}^{-1} \leq \mathbf{g}_1^{-1}\mathbf{I}$ . In comparison in the absolute block operator residual equation (2.2) it only holds that  $0 \leq \mathbf{B}_{\text{abs}}^{-1}$  and so no lower estimate is obtainable. In the more general case of Theorem 3.3 the bounded operator  $\mathbf{B}_{\text{rel}}$  is indefinite but it is always boundedly invertible. For further discussion of the optimality of the Schur complement approach see Section 4.1.

**Proposition 3.7.** *Let  $\lambda_m < \lambda_{m+1}$  and let  $\mathcal{R}(P)$ ,  $\dim \mathcal{R}(P) = m$  be the test space for the form  $h$  such that  $2\eta_m < 1$  then*

$$\frac{1\mu_1}{2\mu_m} \sum_{i=1}^m \frac{\|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2} \leq \sum_{i=1}^m \frac{\mu_i - \lambda_i}{\mu_i}.$$

*Proof.* Let  $u_i \in \mathcal{R}(P)$ ,  $i = 1, \dots, m$  be the Ritz vectors. The proof follows from

$$\begin{aligned} \sum_{i=1}^m \|\mathbf{H} \frac{1}{\mu_i} u_i - u_i\|_{\mathbf{H}^{-1}}^2 &= \sum_{i=1}^m \left[ (u_i, \mathbf{H}^{-1}u_i) - (u_i, \Xi^{-1}u_i) \right] = \sum_{i=1}^m \left[ (u_i, \mathbf{H}^{-1}u_i) - \frac{1}{\mu_i} \right] \\ &\leq \sum_{i=1}^m \left[ \frac{1}{\lambda_i} - \frac{1}{\mu_i} \right] = \sum_{i=1}^m \frac{\mu_i - \lambda_i}{\lambda_i \mu_i}. \end{aligned}$$

Note that according to (3.15) below, we have  $\sum_{i=1}^m \eta_i^2 \leq \sum_{i=1}^m \frac{\|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2}$ .  $\square$

**Remark 3.8.** The upper estimate in the setting of Proposition 3.7 can be achieved by a repeated application of the trace operator and the estimate (3.15) to the identity (3.9). The estimate is rather technical and we leave it out. However, we emphasize that we can recreate the framework of [12, Proposition 2.3] completely.

**3.2. A relationship with standard  $\mathbf{H}^{-1}$ -norm residual estimates.** This section addresses the issue of the computability of  $\eta_i(P)$  by relating these quantities to the standard  $\mathbf{H}^{-1}$ -norm estimates of the residuals associated to the Ritz vector basis of  $\mathcal{R}(P)$ . The proofs as well as the results are technical and as such can be skipped on the first reading.

Let us now reconsider the identities (2.6) and (3.3) and note that they can be understood as generalized matrix eigenvalue problems. Assume  $u_1, \dots, u_m$  are the Ritz vectors from  $\mathcal{R}(P)$ , then for  $i, j = 1, \dots, m$ , we define the matrices

$$\begin{aligned} \Omega_{ij} &= h(\mathbf{H}^{-1}u_i - \Xi^{-1}u_i, \mathbf{H}^{-1}u_j - \Xi^{-1}u_j) \\ \Psi_{ij} &= (u_i, \mathbf{H}^{-1}u_j). \end{aligned}$$

Relation (2.7) from Lemma 2.1 implies that  $\Omega$  is a positive definite matrix and in particular

$$\begin{aligned} \eta_i^2 &= \lambda_i(\Psi^{-1/2}\Omega\Psi^{-1/2}) \\ \Omega_{ii} &= \|\mathbf{H} \frac{1}{\mu_i} u_i - u_i\|_{\mathbf{H}^{-1}}^2 = \frac{1}{\mu_i^2} \|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2, \quad i = 1, \dots, m \\ D_\mu &\leq \Psi \leq (1 + \mathfrak{D}_l)D_\mu, \end{aligned}$$

where  $D_\mu = \text{diag}(\mu_1^{-1}, \dots, \mu_m^{-1})$  and  $\mathfrak{D}_l = \|D_\mu^{-1/2}(\Psi - D_\mu)D_\mu^{-1/2}\|$ . Now, with the help of  $\mu_i = \|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2$ , we obtain

$$\sum_{i=1}^m \lambda_i(D_\mu^{-1/2}\Omega D_\mu^{-1/2}) = \text{tr}(D_\mu^{-1/2}\Omega D_\mu^{-1/2}) = \sum_{i=1}^m \frac{\|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2},$$

and so we conclude that

$$(3.14) \quad \frac{1}{1 + \mathfrak{D}_l} \sum_{i=1}^m \Omega_{ii} \mu_i \leq \sum_{i=1}^m \eta_i^2 \leq \sum_{i=1}^m \Omega_{ii} \mu_i.$$

Estimate (3.14) can now be written as (cf. Proposition 3.7)

$$(3.15) \quad \frac{1}{1 + \mathfrak{D}_l} \sum_{i=1}^m \frac{\|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2} \leq \sum_{i=1}^m \eta_i^2 \leq \sum_{i=1}^m \frac{\|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2}.$$

By a similar argument one can conclude that asymptotically (as  $P$  converges to the  $m$ -dimensional spectral subspace) we have as a heuristic

$$\frac{\|\mathbf{H}u_i - \mu_i u_i\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H}u_i\|_{\mathbf{H}^{-1}}^2} \sim \eta_{m-i+1}^2(P).$$

This indicates that  $\eta_i(P)$  represent a canonical choice of residuals from  $\mathbf{R}(P)$ , i.e. not defined by the Ritz vectors  $u_i$  but rather the vectors which are selected by the variational formulae (3.3).

**Remark 3.9.** The definition of  $\eta_i$  indicates that the problem of computing (or estimating)  $\eta_i$  requires the solution of the  $m \times m$  positive definite generalized eigenvalue problem. Since  $m$  is the multiplicity of the eigenvalue of interest, the computational cost of the solution of such problem is negligible. The main problem is how to evaluate or estimate the moments  $(u_i, \mathbf{H}^{-1}u_j)$ ,  $i, j = 1, \dots, m$  without actually inverting the operator  $\mathbf{H}^{-1}$ . For some possibilities to do this see [1, Section 3.], [14, Section 5.] or [10, Remark 8].

#### 4. A SIMPLE NON-INHIBITED STIFF PROBLEM

These estimates have been used in [16] to study a class of eigenvalue problems which is given by the family of positive definite forms

$$(4.1) \quad h_\kappa(u, v) = h_b(u, v) + \kappa^2 h_e(u, v), \quad \kappa \text{ large}.$$

The forms  $h_b$  and  $h_e$  are assumed to be symmetric, closed and nonnegative and we further assume that  $h_b + h_e$  is positive definite in  $\mathcal{H}$  and that  $\mathcal{Q}(h_b + h_e)$  is dense in  $\mathcal{H}$ . Family (4.1) can always be considered as a perturbation of  $h_b + h_e$  (after an obvious change of variable  $\kappa$ ) rather than as a perturbation of  $h_b$  and so we assume, without affecting the generality of results, that  $h_b$  is positive definite and densely defined.

A detailed study of the spectral property of the families like (4.1) is beyond the scope of this article and will be reported in subsequent publication. We will now consider a very simple problem of this form, and note that (4.1) motivated the example (1.6). Let  $H_0^1[0, 1]$  and  $H_0^1(\mathbb{R}_+)$ ,  $\mathbb{R}_+ := [0, \infty)$  be the standard Sobolev spaces. We also identify the functions from  $H_0^1[0, 1]$

with their extension by zero to the whole of  $\mathbb{R}_+$  and write  $H_0^1[0, 1] \subset H_0^1(\mathbb{R}_+)$ . Consider the family of positive definite forms

$$(4.2) \quad h_\kappa(u, v) = \int_0^\infty u'v' \, dx + \kappa^2 \int_1^\infty uv \, dx, \quad u, v \in H_0^1(\mathbb{R}_+).$$

By  $\mathbf{H}_\kappa$  we denote the positive definite operator defined by  $h_\kappa$  in (4.2). The operators  $\mathbf{H}_\kappa$  converge in the generalized sense to the operator  $\mathbf{H}_\infty$ , which is defined by the form  $h_\infty(u, v) = \int_0^1 u'v' \, dx$ ,  $u, v \in H_0^1[0, 1]$ . For further details on this convergence see [16] and the references therein. We also formally write  $\mathbf{H}_\kappa = -\partial_{xx} + \kappa^2 \chi_{[1, \infty)}$  and  $\mathbf{H}_\infty = -\partial_{xx}$ . As a test function(s) we chose

$$(4.3) \quad u_q(x) = \begin{cases} \sqrt{2} \sin(k\pi x), & 0 \leq x \leq 1 \\ 0, & 1 \leq x \end{cases}, \quad q \in \mathbb{N}.$$

Note that here  $u_q \in \mathcal{Q}(h_\kappa)$  but  $u_q \notin \mathcal{D}(\mathbf{H}_\kappa)$ . The eigenvalues of the operator  $\mathbf{H}_\kappa$  have to be described implicitly. Let  $\mathbf{H}_\kappa v^\kappa = \lambda^\kappa v^\kappa$ , then  $v^\kappa \in C^1(\mathbb{R}_+)$  is

$$v^\kappa(x) = \begin{cases} \sin(\sqrt{\lambda^\kappa} x), & 0 \leq x \leq 1 \\ \frac{\sin \sqrt{\lambda^\kappa}}{e^{-\sqrt{\kappa^2 - \lambda^\kappa}}} e^{-\sqrt{\kappa^2 - \lambda^\kappa} x}, & 1 \leq x \end{cases}$$

and  $\lambda^\kappa$  is a solution of the equation

$$(4.4) \quad \sqrt{\kappa^2 - \lambda^\kappa} = -\sqrt{\lambda^\kappa} \cot(\sqrt{\lambda^\kappa}).$$

The quotient  $\frac{\lambda_1^\infty - \lambda_1^\kappa}{\lambda_1^\infty}$  can be represented (for  $\kappa \rightarrow \infty$ ) by a convergent Taylor series

$$(4.5) \quad \frac{\lambda_1^\infty - \lambda_1^\kappa}{\lambda_1^\infty} = 2 \frac{1}{\kappa} - 3 \frac{1}{\kappa^2} + 8 \left( \frac{1}{2!} + \frac{1}{4!} \pi^2 \right) \frac{1}{\kappa^3} - 10 \left( \frac{1}{2!} + \frac{4}{4!} \pi^2 \right) \frac{1}{\kappa^4} + \dots$$

We directly compute  $\eta_\kappa^2(u_q) := \frac{2}{3+\kappa}$  and combine it with (4.4) and the first order estimate from (2.10) to obtain

$$(1 - \sqrt{\frac{2}{3+\kappa}}) 4\pi^2 =: D(\kappa) \leq \lambda_2(\mathbf{H}), \quad \kappa \geq 5.$$

Theorems 3.3 and 3.4 now yield

$$(4.6) \quad \frac{2}{3+\kappa} \leq \frac{\lambda_1^\infty - \lambda_1^\kappa}{\lambda_1^\infty} \leq \frac{D(\kappa) + \pi^2}{D(\kappa) - \pi^2} \frac{2}{3+\kappa} = \frac{10}{3\kappa} + \frac{1}{\sqrt{\kappa}} O\left(\frac{1}{\kappa}\right), \quad \kappa \geq 5,$$

which is a tight estimate on the behavior of  $\frac{\lambda_1^\infty - \lambda_1^\kappa}{\lambda_1^\infty}$ . Similar estimates hold for other eigenvalues and eigenvectors, too. This example illustrates the “efficiency” of this *a posteriori* estimator. Furthermore, it indicates a role which is played by the first order estimates from [15] in the general theory. For some details of the computation see [16]. The Schrödinger operators in higher dimensions have also been studied in [16]. The estimate for  $\eta^2(u_q)$  can in this case be computed by a use of the advanced probabilistic techniques from [8] or by a use of the boundary layer techniques from [3] (naturally, under the assumption that the domain is finite).

**4.1. A framework for proving the asymptotic exactness.** Let us go back to Remark 3.6. The conclusion of Proposition 3.7 does not appear to be completely satisfactory. The factor  $\mu_1/\mu_m$  limits its applicability to a couple of the lowermost eigenvalues of  $\mathbf{H}$ . The true power of the Schur complement technique can be seen if we rewrite (3.9) as

$$(4.7) \quad \mathbf{I} - \lambda_q \Xi^{-1} = K_s^* K_s + \lambda_q K_s^* \mathbf{W}^{-1/2} (\mathbf{I} - \lambda_q \mathbf{W}^{-1})^{-1} \mathbf{W}^{-1/2} K_s.$$

After applying the trace operator on (4.7) and utilizing Lemma 2.1 we obtain

$$\frac{\sum_{i=1}^m \frac{\mu_i - \lambda_q}{\mu_i}}{\sum_{i=1}^m \eta_i^2} = 1 + \frac{\text{tr}(\lambda_q K_s^* \mathbf{W}^{-1/2} (\mathbf{I} - \lambda_q \mathbf{W}^{-1})^{-1} \mathbf{W}^{-1/2} K_s)}{\sum_{i=1}^m \eta_i^2}.$$

In our  $\kappa$  dependent problem we use this to prove (cf. (4.5) and (4.6))

$$\lim_{\kappa \rightarrow \infty} \frac{\frac{\lambda_q^\infty - \lambda_q^\kappa}{\lambda_q^\infty}}{\eta_\kappa^2(u_q)} = 1.$$

Furthermore, we see why this convergence is pretty rapid. In a general situation we perform this analysis by comparing the singular values  $s_i(K_s(\kappa))$  with  $s_i(\mathbf{W}_\kappa^{-1} K_s(\kappa))$  and noticing that  $s_i(\mathbf{W}_\kappa^{-1} K_s(\kappa))$  is of higher order in  $\kappa^{-1}$ . Here we have assumed an obvious modification of the block matrix representation (3.8) for the  $\kappa$  dependent problem. Exploiting (4.7) in the general setting of (4.1) as well in as in the setting of finite element approximations is beyond the scope of this paper and is a subject of subsequent reports.

## 5. FINITE ELEMENT COMPUTATIONS

As a further explicitly solvable model example let us consider the family of eigenvalue problems

$$(5.1) \quad \begin{aligned} -\psi'' - \alpha \psi &= \omega \psi, \\ e^{i\theta} \psi(0) &= \psi(2\pi), \\ e^{i\theta} \psi'(0) &= \psi'(2\pi), \end{aligned}$$

where  $\theta \in [0, \pi]$  and we chose  $\alpha \in \mathbb{R}$  so that the eigenvalues remain positive. The weak formulation of (5.1) is given by

$$h(\psi, v_i) = \lambda_i(\mathbf{H})(\psi, v_i), \quad \psi \in \mathcal{Q}(h),$$

where

$$(5.2) \quad h(\psi, \phi) := \int_0^{2\pi} (\overline{\psi'} \phi' - \alpha \overline{\psi} \phi), \quad \mathcal{Q}(h) := \{\psi \mid \psi, \psi' \in \mathcal{H}, e^{i\theta} \psi(0) = \psi(2\pi)\}$$

and  $\mathcal{H} = L^2[0, 2\pi]$ . The eigenvalues of the problem (5.1) as well as the Green function of the operator  $\mathbf{H}$ , which is defined by (5.2) are explicitly known, see [28, Theorem XIII.89, Volume 4. pp. 293] and [28, Equation (XIII.154), pp. 292]. In particular we have

$$(5.3) \quad \lambda_1(\mathbf{H}) = \left(-1 + \frac{\theta}{2\pi}\right)^2 - \alpha, \quad \lambda_2(\mathbf{H}) = \left(\frac{\theta}{2\pi}\right)^2 - \alpha, \quad \lambda_3(\mathbf{H}) = \left(1 + \frac{\theta}{2\pi}\right)^2 - \alpha,$$

$$(5.4) \quad v_1(t) = e^{-i\left(-1 + \frac{\theta}{2\pi}\right)t}, \quad v_2(t) = e^{-i\left(\frac{\theta}{2\pi}\right)t}, \quad v_3(t) = e^{-i\left(1 + \frac{\theta}{2\pi}\right)t}$$

N	estimate (3.10)	$\ \mathbf{I} - \lambda \Xi_{P(\mathcal{V}_N^1)}^{-1}\ _H$	estimate (3.7)
40	7.9540e-001	7.9540e-001	7.9558e-001
60	5.1413e-001	5.1413e-001	5.1422e-001
80	3.4389e-001	3.4389e-001	3.4393e-001
100	2.4120e-001	2.4120e-001	2.4123e-001
120	1.7671e-001	1.7671e-001	1.7673e-001

TABLE 1. The performance of the estimates (3.10) and (3.7) on the family of test spaces  $R(P(\mathcal{V}_N^1))$  and for the choice of the norm  $\|\cdot\| = \|\cdot\|_{HS}$ . The computational details can be found in [16, Section 2.7.3, pp. 64].

and<sup>8</sup>

$$(\psi, \mathbf{H}^{-1}\phi) = \int_0^{2\pi} dt_1 \int_0^{2\pi} G(t_1 - t_2) \overline{\psi(t_2)} \phi(t_1) dt_2$$

$$G(t_1 - t_2) = \frac{i}{2\sqrt{\alpha}} \left( e^{i\sqrt{\alpha}|t_1 - t_2|} + \frac{e^{i(t_1 - t_2)\sqrt{\alpha}}}{-1 + e^{-2i\pi\sqrt{\alpha}-i\theta}} + \frac{e^{i(t_2 - t_1)\sqrt{\alpha}}}{-1 + e^{-2i\pi\sqrt{\alpha}+i\theta}} \right).$$

Let us now choose  $\theta = \pi$  and  $\alpha = 0.2499$  for our numerical experiment. With this choice of parameter the problem (5.1) is almost singular and  $\lambda_1(\mathbf{H}) = \lambda_2(\mathbf{H})$ . For  $N \in \mathbb{N}$  define the finite element space

$$\mathcal{V}_N^1 = \{\psi \mid \psi \in C[0, 2\pi], -\psi(0) = \psi(2\pi), \psi \text{ is linear in } \mathcal{I}_p, p = 1, \dots, N\},$$

where  $\mathcal{I}_p := \left\langle \frac{(p-1)2\pi}{N}, \frac{p2\pi}{N} \right\rangle$ , and use

$$(5.5) \quad \mu_i(\mathcal{V}_N^1) := \max_{\substack{\mathcal{S} \subset \mathcal{V}_N^1 \\ \dim \mathcal{S} = \dim \mathcal{V}_N^1 - i}} \min_{\psi \in \mathcal{S} \setminus \{0\}} \frac{h(\psi, \psi)}{(\psi, \psi)}, \quad i = 1, 2$$

to define the Rayleigh-Ritz approximations to the eigenvalue  $\lambda_1(\mathbf{H}) = \lambda_2(\mathbf{H})$ . Let also  $u_i(\mathcal{V}_N^1) \in \mathcal{V}_N^1$ ,  $i = 1, 2$  be two vectors of norm one for which  $\mu_i(\mathcal{V}_N^1) = h[u_i(\mathcal{V}_N^1)]$ ,  $i = 1, 2$  holds. Now, let  $P(\mathcal{V}_N^1)$  be an orthogonal projection onto the linear span of  $\{u_1(\mathcal{V}_N^1), u_2(\mathcal{V}_N^1)\}$  and set  $\Xi_{P(\mathcal{V}_N^1)} = \mathbf{H}_{P(\mathcal{V}_N^1)} P(\mathcal{V}_N^1)$ . We now apply Theorems 3.3 and 3.4 on the projections  $P(\mathcal{V}_N^1)$  and display the results on Table 1.

**5.1. Hierarchical error estimation.** The results from Table 1 show that  $\eta_i, \dots, \eta_m$  accurately capture the behavior of the relative error as  $\frac{1}{N} \rightarrow 0$ . The explicit knowledge of the Green function is most certainly an information which cannot in general be assumed when considering higher dimensional eigenvalue problems. Let us now consider an application of these estimates in the context of the adaptive finite element methods for divergence type elliptic self-adjoint

<sup>8</sup>We implicitly assume that  $\mathcal{H} = L^2[0, 2\pi]$ .

operators in dimension two. We only present a feasibility argument, an algorithmic development will be a subject of a subsequent report.

For the sake of definiteness let  $\mathcal{H} = L^2(\mathcal{R})$ , where  $\mathcal{R}$  is assumed to be a bounded *polygonal domain* and let

$$(5.6) \quad h(u, v) = \int_{\mathcal{R}} (\nabla u)^* \nabla v, \quad u, v \in \mathcal{Q}(h) = H_0^1(\mathcal{R}).$$

By  $H^1(\mathcal{R}) = \{u \in \mathcal{H} : \|u\|^2 + \|\nabla u\|^2 < \infty\}$  we denote the standard first order *Sobolev space*. The gradient  $\nabla$  is meant in the weak sense and  $\|\cdot\|$  denotes the norm on  $L^2(\mathcal{R})$  and  $H_0^1(\mathcal{R}) \subset H^1(\mathcal{R})$  is assumed to be equipped with the norm  $\|u\|_E := \|\nabla u\| = h[u]^{1/2}$  and it consists of those  $H^1(\mathcal{R})$  functions which vanish on the boundary of  $\mathcal{R}$  in the sense of the trace operator.

The set  $\mathcal{T}_d$  is called a *triangulation of the polygonal domain  $\mathcal{R}$*  if it consists of the triangles such that union of these triangles is  $\overline{\mathcal{R}}$  and such that the intersection of two such triangles either consists of a common side or of a common vertex of both triangles or is empty. By  $d$  we denote the maximal diameter of all triangles in  $\mathcal{T}_d$ . For a given triangulation  $\mathcal{T}_d$  we define the finite dimensional function spaces:

$$\begin{aligned} \mathcal{V}_d^1 &= \{u \in \mathcal{Q} \mid u \in C(\overline{\mathcal{R}}) \text{ and } v|_K \text{ is linear, } K \in \mathcal{T}_d\} \\ \mathcal{V}_d^2 &= \{u \in \mathcal{Q} \mid u \in C(\overline{\mathcal{R}}) \text{ and } v|_K \text{ is quadratic, } K \in \mathcal{T}_d\} \end{aligned}$$

and the orthogonal projections  $V_{i,d}$ ,  $i = 1, 2$  such that  $\mathbf{R}(V_{i,d}) = \mathcal{V}_d^i$ ,  $i = 1, 2$ . To simplify the notation we write  $\mathbf{H}_{i,d} := \mathbf{H}_{V_{i,d}}$ ,  $i = 1, 2$  and also define the orthogonal projection  $P_d$  such that  $\mathbf{R}(P_d)$  equals the linear span of  $\{u_{1,d}, u_{2,d}\}$ . In what follows we assume, as in [9], that  $\mathcal{T}_d$  is graded and shape regular family of triangulations and that it satisfies the nondegeneracy property [9, Assumption 4.1]. Let us assume that we have

$$(5.7) \quad \mu_{i,d} := \max \left\{ \min_{\psi \in \mathcal{S} \setminus \{0\}} \frac{h(\psi, \psi)}{(\psi, \psi)} : \mathcal{S} \subset \mathcal{V}_d^1, \dim \mathcal{S} = \dim \mathcal{V}_d^1 - (i+1) \right\}, \quad i = 1, 2$$

and  $u_{i,d} \in \mathcal{V}_{T_d}^1$ ,  $i = 1, 2$  are chosen so that<sup>9</sup>  $\mathbf{H}_{1,d} u_{i,d} = \mu_{i,d} u_{i,d}$ ,  $i = 1, 2$ . The result [9, Theorem 1.1] and in particular the last remark on [9, pp. 12] yield the estimate

$$(5.8) \quad \frac{h[\mathbf{H}^{-1} u_{i,d} - \mathbf{H}_{2,d}^{-1} u_{i,d}]^{1/2}}{h[\mathbf{H}^{-1} u_{i,d} - \mathbf{H}_{1,d}^{-1} u_{i,d}]^{1/2}} = \frac{\|\mathbf{H}^{-1} u_{i,d} - \mathbf{H}_{2,d}^{-1} u_{i,d}\|_E}{\frac{1}{\mu_{i,d}} \|\mathbf{H} u_{i,d} - \mu_{i,d} u_{i,d}\|_{\mathbf{H}^{-1}}} \leq \alpha, \quad i = 1, 2$$

with the constant  $\alpha$  which depends solely on the shape regularity of  $\mathcal{T}_d$ . Set  $r_{i,d} = \mathbf{H}_{2,d} u_{i,d} - \mu_{i,d} u_{i,d}$ . Combining (5.8) and [2, Estimate (2.16)] we conclude that there exists constants  $C_*$  and  $c_*$ , solely depending on the shape regularity of  $\mathcal{T}_d$ , such that

$$(5.9) \quad c_* \frac{\|r_{i,d}\|_{\mathbf{H}_{2,d}^{-1}}^2}{\mu_{i,d}} \leq \frac{\|\mathbf{H} u_{i,d} - \mu_{i,d} u_{i,d}\|_{\mathbf{H}^{-1}}^2}{\|\mathbf{H} u_{i,d}\|_{\mathbf{H}^{-1}}^2} \leq C_* \frac{\|r_{i,d}\|_{\mathbf{H}_{2,d}^{-1}}^2}{\mu_{i,d}}, \quad i = 1, 2.$$

This estimate can now be directly plugged into the trace type estimates from Theorem 3.4 or Proposition 3.7. Furthermore, Remark 3.9 allows us to exploit other unitary invariant norms with similar ease.

<sup>9</sup>This can be checked by a direct computation.

**Remark 5.1.** Note that  $\mathbf{H}_{2,d}|_{\mathcal{V}_d^2} \in \mathcal{L}(\mathcal{V}_d^2)$ . The  $\mathbf{H}_{2,d}^{-1}$  norms of the residuals  $r_{i,d} \in \mathcal{V}_{T_d}^2$  can efficiently (cheaper than when solving a linear system) be approximated as functions of the vectors  $w_{i,d} \in \mathcal{V}_{T_d}^2 \ominus \mathcal{V}_{T_d}^1$ ,  $i = 1, 2$ , as given by [2, Theorem 2.1]. Similar consideration has been explored in [25, Estimates (29)–(30)] (cf. [2, Theorem 2.2]), but in comparison our estimates give more explicit information on the dependence of the constants on the mesh and provide the optimality argument, too.

To summarize, the arguments of Remark 3.9 indicate that it is possible to estimate the  $\mathbf{H}^{-1}$  norm of the residual cheaper than it takes to solve the linear system. Furthermore, we have shown that when deciding on the convergence of the finite element method the size of  $\|r_{i,d}\|_{\mathbf{H}_{2,d}^{-1}}^2 / \mu_{i,d}$  should be compared to the relative gap measure  $\mathbf{g}_q$  to decide if the approximation is good enough. By this we mean if the whole multiplicity of the target eigenvalue has been resolved by  $\mathbf{R}(P_d)$ . To get a feeling for this statement one should remember the picture of [12, Example 2.1].

## 6. CONCLUSION

The main benefit of our approach is that, as the theoretical considerations from Section 4 and Table 1 corroborate, up to (5.9) we have had globally optimal estimates for the eigenvalue error (i.e. almost no information was lost). After (5.9) we have started aggressively trading off accuracy for speed. Our theory is such that this can be achieved, in numerous situations, by a simple combination of the Galerkin orthogonality condition and any of the “of the shelf” results like the those from [2, 9] or [3, 8] in the singularly perturbed setting. On top of this comes the heuristic insight from [12, Section 3] which indicates that we have properly identified the components of the error as given by (1.2). It should be noted that our Theorem 3.3 directly corresponds to [12, Proposition 2.3], since both are motivated by [11]. Furthermore, the estimates from Theorems 3.3 and 3.4 can be combined with [15, Theorem 6.1] and the well known identity from [23, Ad (v), pp.617] to obtain optimal estimates for the eigenvalue error  $\|u_i - v_i\|_E / \|v_i\|_E$ . The estimates can even be obtained in a situation in which the multiple eigenvalue splits in a cluster of eigenvalues. This is a subject of the followup report. As a conclusion let us remember the remarks 1), 2), 3) from the Introduction. We have introduced a matrix analytic techniques which tackle both test vectors outside the domain of definition of the operator and the multiplicity of the approximated eigenvalue in a natural and constructive way. Furthermore, Remark 3.9 opens a way to exploiting other unitary invariant norms for scaling robust eigenvalue estimation.

## ACKNOWLEDGEMENT

The author would like to thank Prof. Dr. Krešimir Veselić, Hagen and Prof. Dr. Volker Enss, Aachen for helpful discussions and support during the research and the preparation of this manuscript. The author also thanks Dr. Mario Arioli, Didcot for a helpful discussion and for pointing out the reference [14].

## REFERENCES

- [1] M. Arioli. A stopping criterion for the conjugate gradient algorithms in a finite element method framework. *Numer. Math.*, 97(1):1–24, 2004.
- [2] R. E. Bank and R. K. Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Numer. Anal.*, 30(4):921–935, 1993.
- [3] V. Bruneau and G. Carbou. Spectral asymptotic in the large coupling limit. *Asymptot. Anal.*, 29(2):91–113, 2002.
- [4] F. Chatelin. *Spectral approximation of linear operators*. Academic Press Inc., 1983.
- [5] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970.
- [6] C. Davis, W. M. Kahan, and H. F. Weinberger. Norm-preserving dilations and their applications to optimal error bounds. *SIAM J. Numer. Anal.*, 19(3):445–469, 1982.
- [7] J. Demmel and K. Veselić. Jacobi’s method is more accurate than QR. *SIAM J. Matrix Anal. Appl.*, 13(4):1204–1245, 1992.
- [8] M. Demuth, F. Jeske, and W. Kirsch. Rate of convergence for large coupling limits by Brownian motion. *Ann. Inst. H. Poincaré Phys. Théor.*, 59(3):327–355, 1993.
- [9] W. Dörfler and R. H. Nochetto. Small data oscillation implies the saturation assumption. *Numer. Math.*, 91(1):1–12, 2002.
- [10] Z. Drmač. On relative residual bounds for the eigenvalues of a Hermitian matrix. *Linear Algebra Appl.*, 244:155–163, 1996.
- [11] Z. Drmač and V. Hari. Relative residual bounds for the eigenvalues of a Hermitian semidefinite matrix. *SIAM J. Matrix Anal. Appl.*, 18(1):21–29, 1997.
- [12] Z. Drmač and K. Veselić. New fast and accurate Jacobi SVD algorithm: II. *to appear in SIAM Journal on Matrix Analysis and Applications*. LAPACK Working Note 170.
- [13] R. G. Durán, C. Padra, and R. Rodríguez. A posteriori error estimates for the finite element approximation of eigenvalue problems. *Math. Mod. Meth. Appl. Sci.*, 13(8):1219–1229, 2003.
- [14] G. H. Golub and G. Meurant. Matrices, moments and quadrature. In *Numerical analysis 1993 (Dundee, 1993)*, pages 105–156. Longman Sci. Tech., Harlow, 1994.
- [15] L. Grubišić. On eigenvalue estimates for nonnegative operators. *SIAM J. Matrix Anal. Appl.*, 28(4):1097–1125, 2006.
- [16] L. Grubišić. *Ritz value estimates and applications in Mathematical Physics*. PhD thesis, Fernuniversität in Hagen, dissertation.de Verlag im Internet, ISBN: 3-89825-998-6, 2005.
- [17] L. Grubišić and K. Veselić. On Ritz approximations for positive definite operators I (theory). *Linear Algebra and its Applications*, 417(2-3):397–422, 2006.
- [18] V. Heuveline and R. Rannacher. A posteriori error control for finite approximations of elliptic eigenvalue problems. *Adv. Comput. Math.*, 15(1-4):107–138 (2002), 2001.
- [19] I. C. F. Ipsen. A note on unifying absolute and relative perturbation bounds. *Linear Algebra Appl.*, 358:239–253, 2003.
- [20] T. Kato. On the upper and lower bounds of eigenvalues. *J. Phys. Soc. Japan*, 4:334–339, 1949.
- [21] T. Kato. *Perturbation theory for linear operators*. Springer-Verlag, 1976.
- [22] M. Kraus, M. Langer, and C. Tretter. Variational principles and eigenvalue estimates for unbounded block operator matrices and applications. *J. Comput. Appl. Math.*, 171(1-2):311–334, 2004.
- [23] M. G. Larson. A posteriori and a priori error analysis for finite element approximations of self-adjoint elliptic eigenvalue problems. *SIAM J. Numer. Anal.*, 38(2):608–625, 2000.
- [24] R.-C. Li. Relative perturbation theory. I. Eigenvalue and singular value variations. *SIAM J. Matrix Anal. Appl.*, 19(4):956–982, 1998.
- [25] K. Neymeyr. A posteriori error estimation for elliptic eigenproblems. *Numer. Linear Algebra Appl.*, 9(4):263–279, 2002.
- [26] E. Ovtchinnikov. Cluster robust error estimates for the Rayleigh–Ritz approximation II: Estimates for eigenvalues. *Linear Algebra Appl.*, 415(1), 2006.

- [27] B. N. Parlett. *The symmetric eigenvalue problem*. Prentice-Hall Inc., 1980.
- [28] M. Reed and B. Simon. *Methods of modern mathematical physics. I–IV*. Academic Press, 1978.
- [29] E. Sánchez-Palencia. Asymptotic and spectral properties of a class of singular-stiff problems. *J. Math. Pures Appl.* (9), 71(5):379–406, 1992.
- [30] B. Simon. *Trace ideals and their applications*. Cambridge University Press, 1979.
- [31] J.-G. Sun, Eigenvalues of Rayleigh quotient matrices. *Numer. Math.*, 59:603–614, 1991.
- [32] G. Temple. The theory of Rayleigh’s principle as applied to continuous systems. *Proc. Roy. Soc. London. Ser. A.*, 119:276–293, 1928.

INSTITUT FÜR REINE UND ANGEWANDTE MATHEMATIK, RWTH AACHEN, TEMPLERGRABEN 55, D-52056 AACHEN, GERMANY. (ON LEAVE FROM PMF-DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ZAGREB, CROATIA)

*E-mail address:* luka.grubisic@iram.rwth-aachen.de